Scientometrics (2018) 115:1253–1269
https://doi.org/10.1007/s11192-018-2727-5

#Psychology: a bibliometric analysis of psychological literature in the online media

Sebastian Vogl[1] · Thomas Scherndl[1] · Anton Kühberger[1,2]

Received: 2 October 2017 / Published online: 31 March 2018
© The Author(s) 2018

Abstract Online media and especially social media are becoming more and more relevant to our everyday life. Reflecting this tendency in the scientific community, alternative metrics for measuring scholarly impact on the web are increasingly proposed, extending (or even replacing) traditional metrics (e.g., citations, journal impact factor, etc.). This paper explores the relationship between traditional metrics and alternative metrics for psychological research in the years from 2010 to 2012. Traditional publication metrics (e.g., number of citations, impact factor) and alternative metrics (collected from *Altmetric*, a website that collects and counts references as they appear in Wikipedia, public policy documents, research blogs, mainstream media, or social networks) were extracted and compared, using a dataset of over 245,000 publications from the Web of Science. Results show positive, small to medium, correlations on the level of individual publications, and frequently medium to high correlations on the level of research fields of Psychology. The more accumulated the level of analysis, the higher the correlations. These findings are fairly robust over time and comparable to findings from research areas other than Psychology. Additionally, a new metric, the *Score Factor*, is proposed as a useful alternative metric to assess a journal's impact in the online media.

Keywords Altmetrics · Online media · Psychology · Scholarly impact · Social media

✉ Anton Kühberger
anton.kuehberger@sbg.ac.at

[1] Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria

[2] Centre of Cognitive Neurosciences, University of Salzburg, Salzburg, Austria

Introduction

Ever since the dawn of the Information Age (Castells 1999), data are collected and spread rapidly online. This is especially true for social media, such as Facebook or Twitter, where information gets distributed swiftly. Online media becomes more and more relevant to our daily life, as a 2016 report from the Pew Research Center shows: 66% of Facebook and 59% of Twitter users (47 and 52% in 2013, respectively) are getting news from their social networking site. Because more or less every researcher nowadays is searching for scientific information on the internet, standard databases like *PubMed*, but also online media have gained enormous impact on dissemination of scientific work (Brossard and Scheufele 2013). This development opens up a new approach of measuring scientific influence and puts traditional measures of scholarly success into question.

The measurement of outstanding achievement in Science has a long tradition. More than 90 years ago, Lotka (1926) published his famous scientometric formula, known as *Lotka's Inverse Square Law of Scientific Productivity*. Based on the investigation of the name indexes of standard reference tools at the time (*Chemical Abstracts*, and *Auerbach's Geschichtstafeln der Physik*) he proposed that the relationship between that the number of scientists making at least one contribution (x) and the frequency of their contributions (y) is constant: $x^n y$ = const., with $N = 2$ (therefore the name square law). Specifically, in any set of authors, about 60% make one single contribution. In addition, if 100 authors were contributing one paper each, 25 would be contributing two papers each ($1/2^2$, i.e., 25%), 11 would be contributing three papers each ($1/2^3$, i.e., 11.1%), the number of authors contributing four papers each would be about 6 ($1/2^4$, i.e., 6%), and so on. Thus, the number of researchers making n contributions is about $1/2^n$. Lotka's law was an approximation for the data he had at hand in 1926. It is still more off in describing publication productivity in more recent years (Coile 1977). It thus is in need of adjustment (Nath and Jackson 1991), but it still informs thinking about the measurement of scientific productivity. For instance, it implies that it is a small percentage of researchers who are responsible for the lion's share of the work. Dennis (1954) surveyed about 80 years of research in psychology and accordingly found that the top 10% produced about 50% of the publications, and the less productive half contributed 15% or less.

Another classic scientometric rule exists between the quantity and the quality of research output, such that researchers who are most productive also are, on average, most creative (the *constant-probability-of-success model*; Simonton 1988a, b). This implies that the number of citation a researcher receives is a positive function of his or her total number of publications (Rushton 1984). Interestingly, total productivity is also closely related to number of citations of the three best publications (Cole and Cole 1973).

Rushton (1984) was positive about citation counts, and indeed the now ubiquitous Science Citation Index, created in the 1960s (Garfield 1964) became a central measure of scholarly work in academics (Smith and Fiedler 1970). Thompson Reuter's yearly Journal Citation Report is one of the largest reports on research influence on the journal and category level, using citations. Measures of scientific publishing are still being developed (e.g., the *TRank* measure by Zhang et al. 2017), and the influence of indexes like the *Social Science Citation Index* (S)SCI is still growing, but the growth rate of publication using new channels, like conference proceedings, open archives, blogs, and home pages exceeds that of the traditional channels. This declining coverage in SCI and especially in SSCI is problematic (Larsen and von Ins 2010). Thus, in recent years additional indices have been appearing. These measure the immediate rather than long run impact of scholarly work, not only in academia, but also in popular media, thus tapping a different source of information

for evaluating scientific impact. The question is whether these alternative metrics—as implied by the name—really are alternative to traditional metrics of scholarly impact.

Alternative metrics to the traditional scientific metrics, measuring the impact of research on the web, are called *Altmetrics*, following a proposal by Priem and Hemminger (2010). Altmetrics (similar to *Webometric*; Almind and Ingwersen 1997; Thelwall et al. 2005) collect bibliometric, scientometric, and informetric data on the World Wide Web. Thus they provide access to various types of information pertaining to scholarly publications, most notably on coverage, density, and intensity. We use these terms as defined by Haustein et al. (2015; p. 5): "Coverage is defined as the percentage of papers with at least one social media event or citation. Density is the average number of social media counts or citations per paper (i.e., considering all publications included in the study), while intensity indicates the average number of social media or citation counts for all documents with at least one event (non-zero counts)." These three measures provide different perspectives on the literature. Coverage indicates the chances for being included in the social media market, presumably being influenced by fads and fashions in science. Density and intensity do provide partly overlapping perspectives as they are correlated, since intensity is measured for the subset of documents with at least one count. We expect the distribution of density to be concentrated at the level of zero, since most documents fail in getting any attention in the social media. Less can be said about the expected distribution of intensity: it depends on how the score is calculated (see below). Note, however, that the notions of coverage, density, and intensity are not used consistently in the literature. The ambiguous use of these measures is problematic and may be one of the reasons why some authors warn of altmetrics as a dangerous idea, especially if it is used for measuring the quality of research or a researcher (e.g., Colquhoun and Plested 2014; Gumpenberger et al. 2016).

Coverage of documents varies with source, and *Twitter* is the platform providing the highest coverage (Thelwall et al. 2013). Twitter fares well compared to other social web services especially when it comes to science topics. Eysenbach (2011) even went as far as proposing a "twimpact factor" to measure research uptake on Twitter. Various disciplines such as astrophysics (Haustein et al. 2014a) or biomedicine (Haustein et al. 2014b), as well as journals such as PLoS ONE (de Winter 2015) have already been analyzed on their presence on Twitter. Interestingly, correlations between tweets and citations generally were found to be low (Patthi, et al. 2017), implying a difference between impact metrics based on tweets and those based on citations. Indeed, most research suggests that there is little (or moderate at best) relationship between citations and altmetrics for Twitter, as well as for other platforms such as *Mendeley* (Zahedi et al. 2014; Bar-Ilan et al. 2012), and over all platforms (Costas et al. 2015). To the best of our knowledge, no comparable research exists on the relationship between traditional and altmetrics for the psychological literature. This is reported here. We try to provide a comprehensive picture of the coverage, density, and intensity of psychological research in altmetrics from all over the web, rather than focusing on a single online media platform. This can be done by using something called the *Altmetric Score*. This score accumulates hits over all altmetrics data types and it can be gathered from *Altmetric.com*. Altmetric.com is a website that is dedicated to collecting all kinds of altmetrics data for calculating a sum score, the Altmetric Score (AS). The AS expresses the weighted amount of traffic that some publication or research generates on the web. It uses three main factors. (i) volume—measured by the number of people mentioning a paper; (ii) source where the piece is mentioned, with sources weighted differently; and (iii) authors—a count of who mentions something to whom.

The AS has limitations, most notably with respect to transparency, standardization, and consistency (Gumpenberger et al. 2016). However, it is the best measure available to tap

various different sorts of activities in the social media. We collected the AS and compared it to traditional scores of scientific impact, to investigate the relationships between scientific fame and popular fame. Haustein et al. (2015) did something similar, but they concentrated on a single year (2012), a broad categorization of fields, and on document types. We concentrate on psychological research, as indexed by publications in the period from 2010 to 2012. Thus, all papers related to Psychology published between 2010 and 2012, identified by a unique digital identifier, constitute our sample. For this sample of papers we extract metrics on four levels: field, journal, article, and source. We will provide analyses of these four different levels of aggregation.

(i) Field analysis. This investigates the AS and citation scores for various fields and subfields of Psychology. We hope to identify fields and subfields that are especially popular in the online media. We expect that (sub)fields differ with respect to popularity, and that citation popularity is relatively unrelated to popularity as measured with altmetrics.
(ii) Journal analysis. This analysis identifies journals that have the most impact in online media and investigates the correlation between traditional metrics and altmetrics at the level of journals. We expect to replicate that this correlation generally is low, around $r = .20$.
(iii) Article analysis. This analysis measures the relationship between the AS and article impact metrics for individual articles. In a focused analysis we identify the ten highest scoring articles.
(iv) Source analysis. This analysis identifies the online sources that are the most receptive for psychological articles and investigates the relationship between the AS and citation counts for each source. In line with Thelwall et al.(2013) we expect Twitter to be the most important source.

Method

Measures

To provide for a broad picture of visibility in terms of altmetrics of the psychological literature, we extracted the *Altmetric Score* (AS). The AS is calculated by *Altmetric.com,* a company that specializes in tracking and quantifying the coverage, density, and intensity of content in different alternative sources. It includes a number of different sources, with sources weighted by the likelihood of online sharing. Thus, the weighting reflects the source's potential impact on the online society. Specifically, news (number of times a paper appears in a news outlet online, such as *ZEIT Online* or *Forbes*) gets the highest weight (w = 8.00), followed by blogs (frequency of appearance in a blog; $w = 5.00$), *Twitter* ($w = 1.00$), *Google*$^+$ ($w = 1.00$), and *Facebook* ($w = .25$). All other sources (e.g., *Wikipedia*, *Reddit*, *LinkedIn*) are merged into one variable named *Other,* with weightings between .25 and 3.00 (see https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-. For getting an impression the reader can download an app from Altmetric.com called *Altmetric it!* to get the AS for individual papers and learn about its sources). Although the AS has some qualitative components to it, it is not a measure for the excellence of a researcher's work, but only indicates a papers' online attention.

In addition to the AS, we calculate a new measure, called the *Score Factor* (SF), to compare journals with regard to their influence in alternative online media. The basic problem is that the AS is incomplete as an index of a journal's alternative impact, since it contains only the papers that have an AS greater than zero, i.e., those that are covered in some online source. However, the problem is that most papers do not make it into any alternative metric, and thus their AS = 0. In contrast, citation coverage is much higher than coverage in any of the social media metrics. For instance, Haustein et al. (2015) report an average citation rate of 3.17, but an average Twitter coverage of only .78, although Twitter has by far the most coverage in social media. Our SF takes this into account by using two different scores acquired from altmetrics: The percentage of all papers which have been scored (AS > 0) for a certain journal ($P^{\%}_{\text{Scored}}$); and the mean AS for those papers which have been scored ($M^{\text{AS}}_{\text{Scored}}$). That is, SF is an altmetric score, weighing density by coverage.

$$\text{SF} = P^{\%}_{\text{Scored}} \times M^{\text{AS}}_{\text{Scored}}$$

For a journal to achieve a high SF, a high AS score has to be paired with frequent coverage in the online media.

Data

Data were acquired from the Web of Science (WoS) in June 2016. Eligible papers were articles pertaining to the discipline of Psychology, published between 2010 and 2012. This search resulted in 245,630 single papers. We used the Digital Object Identifier (DOI), or the PubMed-ID for identifying papers, since a DOI (or another unique identifier) is needed for the retrieval of bibliometric information from Altmetric.com. Identifiers were available for 239,910 papers. Papers were matched to fields by using an open-access classification tool acquired from *Science-Metrix.com*. This tool is based on a hierarchical, three-level classification tree and assigns journals to mutually exclusive categories (Archambault et al. 2011). The highest level in this classification is *domain*, including, for instance, Applied Sciences, Arts and Humanities, or Economic and Social Sciences. We did not use this level, as we include only papers from psychology. We did, however, use the next two levels, *field*, and *subfield*. Classification of papers into fields and subfields was possible for 213,738 papers. Journal-level analysis was done only for journals with a Journal Impact Factor. Journal Impact Factors were taken from the Thomson Reuter's 2014 Journal Citation Report. The 2014 report helped to deal with the problem of citation lag, since a journal's impact factor is calculated by the number of citations in the two years to follow the publication year. Journal impact factors were available for 202,432 papers. Finally, we did some data-cleaning by excluding journals that did not reach a minimal count of 20 among the 202,432 papers (i.e., < .01%). The article-level analysis included only papers with AS > 0. These were 57,087 papers, representing a coverage of 28%. Figure 1 displays the selection and classification procedure.

Results

Fields

The 202,432 papers were classified into 21 different fields, containing 125 different subfields. Table 1 displays the 21 fields, and 17 selected subfields. Subfields were selected if

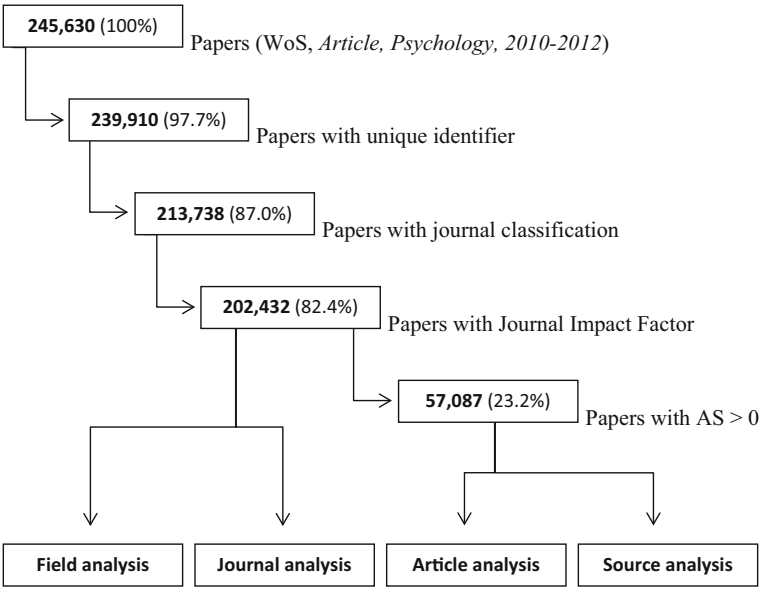

Fig. 1 Data selection tree

they either (i) contained at least 3000 papers, or (ii) were classified into the field *Psychology and Cognitive Sciences*. Note that many papers from Clinical Psychology were classified as *Clinical Medicine*, rendering this by far the most voluminous field (81,762 papers, i.e., 40.4%), considerably larger than *Psychology and Cognitive Sciences* (46,189 papers, i.e., 22.8%), which was the second ranked field in terms of the number of papers published.

Since both citations and AS were positively skewed, a log transformation was applied on the data before doing the analyses.[1] First, we found a strong positive correlation (r_S = .503, p = .020) between mean AS and mean citation frequency of all scored articles for the 21 fields. A similar result was found for the subfields (N = 125; r_S = .417, $p < .001$). The average correlation over all fields (see Table 1, excluding *Built Environment and Design*, since this field had only two articles scored) was $r_{\log/\log}$ = .294, with 16 out of 21 correlations being significant at least at $p < .05$.

In terms of productivity, *Clinical Medicine* is in the lead, with nearly half of all published papers pertaining to this field. *Psychology and Cognitive Sciences* is also a very productive field, as is *Public Health and Health Services*. Note however, that papers were only included if they were related to the discipline of Psychology.[2] The highest scoring

[1] When appropriate we transformed frequencies by first adding 1 to all counts and then taking the logarithm. Adding 1 avoids losing zero counts and retains the zero point (since log(0) is undefined, and log(1) = 0). Transforming the data had negligible effects on the size of the correlations as compared to Spearman correlations (r_s).

[2] Some fields in Table 1 may appear strange (e.g., Physics & Astronomy; Chemistry), but the categorization can be defended. For instance, the article "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion" is published in the *Journal of the Acoustical Society of America*, categorized as Physics and Astronomy. Many papers on acoustics investigating physiological or neural effects of noise are therefore categorized as Physics and Astronomy. Similarly, physiological effects are often categorized as Chemistry.

Table 1 Altmetric Scores and citation frequencies for all fields and selected subfields

Field / Subfield	P	$P^{\%}_{\text{Scored}}$	AS_{Total}	M_{Scored} AS	M_{Scored} Citations	$r_{\text{log/log}}$
Agriculture, Fisheries and Forestry	1391	17.5	1297	5.3	10.3	.212***
Biology	1236	42.0	5836	11.2	22.0	.176***
Biomedical Research	7908	31.7	29,196	11.6	22.6	.326***
Built Environment and Design	30	6.7	13	6.5	18.5	–
Chemistry	815	28.5	1169	5.0	12.6	.110
Clinical Medicine	81,762	25.2	160,016	7.8	21.8	.292***
General and Internal Medicine	6235	31.8	34,477	17.4	18.8	.423***
Neurology and Neurosurgery	20,297	22.8	31,446	6.8	30.0	.362***
Oncology and Carcinogenesis	3282	19.1	3649	5.8	21.6	.318***
Psychiatry	15,911	26.7	30,172	7.1	25.0	.322***
Communication and Textual Studies	185	28.6	232	4.4	7.8	.467**
Earth and Environmental Sciences	297	16.5	614	12.5	21.9	.274
Economics and Business	3259	28.3	5375	5.8	18.9	.250***
Enabling and Strategic Technologies	1082	27.2	2580	8.8	14.9	.387***
Engineering	591	13.4	736	9.3	19.4	.305**
General Arts, Humanities and Social Sciences	259	35.9	628	6.8	7.5	.316**
General Science and Technology	5394	70.7	79,806	20.9	23.1	.344***
Historical Studies	496	20.8	817	7.9	5.5	.111
Information and Communication Technologies	2668	31.6	5615	6.7	15.2	.325***
Mathematics and Statistics	188	17.0	78	2.4	34.7	.231
Philosophy and Theology	1544	21.6	1758	5.3	6.5	.409***
Physics and Astronomy	841	7.6	191	3.0	10.1	.454***
Psychology and Cognitive Sciences	46,189	29.5	116,294	8.5	19.0	.292***
Behavioral Science and Comparative Psychology	3107	26.3	5819	7.1	22.0	.254***
Clinical Psychology	6461	33.7	11,751	5.4	16.4	.257***
Developmental and Child Psychology	7069	31.5	13,134	5.9	19.0	.282***
Experimental Psychology	17,008	28.1	49,259	10.3	21.4	.276***
General Psychology and Cognitive Sciences	772	22.4	686	4.0	12.6	.325***
Human Factors	1653	16.9	954	3.4	10.3	.106
Social Psychology	9414	32.3	34,403	11.3	18.3	.358***
Psychoanalysis	705	17.9	288	2.3	2.9	.174
Public Health and Health Services	38,259	27.7	56,860	5.4	13.4	.301***
Nursing	6910	17.5	3430	2.8	7.9	.266***
Public Health	13,485	35.2	25,585	5.4	13.5	.296***
Rehabilitation	5729	17.2	3368	3.4	12.1	.219***
Substance Abuse	5125	30.5	9238	5.9	15.3	.319***
Social Sciences	8038	26.0	11,638	5.6	13.0	.290***

Table 1 continued

Field / Subfield	P	$P^{\%}_{\text{Scored}}$	AS_{Total}	M_{Scored} AS	M_{Scored} Citations	$r_{\text{log/log}}$
Education	3207	22.9	4527	6.2	16.1	.334***

P = Number of papers; $P^{\%}_{\text{Scored}}$ = Percentage of papers scored (coverage); AS_{Total} = Accumulated AS over all papers for each field or subfield; M_{Scored} = Mean of all papers scored (intensity); $r_{\text{log/log}}$ = correlation between AS and citation frequencies on log-transformed frequencies; frequencies were increased by adding 1

**$p < .01$; ***$p < .001$

field in terms of the AS (see column AS in Table 1) was *General Science and Technology* (M = 20.9, SD = 65.1), with 3811 out of 5394 (70.7%) articles being scored (see column $P^{\%}_{\text{Scored}}$). *Psychology and Cognitive Sciences* is in the middle of the pack. Among the 21 fields it covers rank 6 in percentage of articles scored, 15 in AS, 14 in citations, and 12 in correlation.

Mathematics and Statistics stands out with the highest number of citations per paper scored (M = 34.7, SD = 128.4), and with the lowest AS (M = 2.4, SD = 2.6). Although Haustein et al. (2015) used a different classification and separate scores for different alternative media, the findings do closely match: some topics and fields enjoy greater popularity in the social media, presumably because they represent the "softer" sciences and are easier to understand by the lay audience. Formal content does not lend itself to easy online sharing. In addition, general topics may be particularly interesting for being shared online.

Journals

The 202,432 papers were published in 3644 different journals of which 1838 met the threshold of at least 20 papers. Since the score here is a summary score over journals, coverage is high: most of the 1838 journals have been scored at least once (1591, i.e., 86.6%). *PloS ONE* scored highest, accumulating an AS of 53,597 for 3361 out of 4615 articles (M = 15.9, SD = 56.3). Note the high standard deviation, indicating the long tail that is typical for this type of data. The highest percentage of articles per journal scored was achieved by *Cell*, with 32 out of 34 articles scored (94.1%). *Science* had the highest AS (AS = 65.2, SD = 107.6) per article scored. Figure 2 depicts the relationship between percentage of papers scored per journal and AS per article scored.

Figure 2 shows that, even at the level of journals, the data are heavily skewed, since most journals have a small mean AS, often near zero. Indeed, most journals have an AS fairly below 10, and less than half of their articles are scoring in AS. Some journals are outstanding: for instance, *PloS Medicine,* and *Science* have more than 80% of their articles scored, *Science* with a mean AS > 60, *PloS Medicine* with mean AS > 40. *Nature* has about 50% of their papers scored, with AS > 60. Note, however, that a big impact factor does not automatically guarantee a high AS, since high impact journals exist in all four quadrants of Fig. 2.

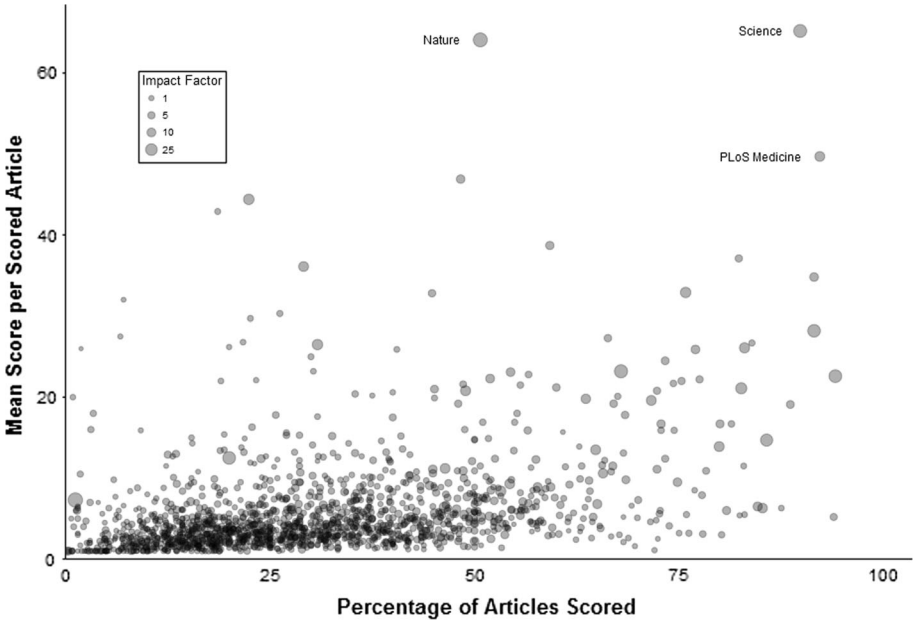

Fig. 2 Scatterplot of all scored journals (N = 1591). Percentage of articles scored by mean AS is plotted. Impact factor of journal is indicated by size of the points. The three journals with the highest Score Factor (see text) are labeled

Spearman correlations between alternative metrics and journal impact factor are shown in Table 2. Also reported is the Scoring Factors (SF), which is the weighted AS (coverage × density). AS, $P^{\%}_{\text{Scored}}$, and IF are correlated in a similar size ($r \approx .40$), indicating that they tap, to a degree, similar information. Interestingly, this correlation is about double the correlation reported in Haustein et al. (2014a, b) for the relationship between Twitter metrics and citations, indicating that the SF is a better predictor of citation impact than tweets. Mendeley seems to be an even better predictor than the SF, correlating around .5 (Zahedi et al. 2014). It is important to bear in mind that these correlations are measured at the level of journals, not of individual papers. That is, among the journals scoring at all, if a journal has a high mean AS, or if it has a high percentage of papers scoring, this journal also tends to have a high impact factor. The high correlation of SF with AS and $P^{\%}_{\text{Scored}}$ is a

Table 2 Spearman correlations between alternative and classic journal metrics

	$M^{\text{AS}}_{\text{Scored}}$	$P^{\%}_{\text{Scored}}$	SF	IF
$M^{\text{AS}}_{\text{Scored}}$	1.000	.444	.840	.401
$P^{\%}_{\text{Scored}}$	.444	1.000	.832	.426
SF	.840	.832	1.000	.477

Based on N = 1591 journals. M^{AS}_{Scored} = Mean score for all scored articles for each journal; $P^{\%}_{\text{Scored}}$ = Percentage of scored papers for each journal; SF Score Factor. All correlations are significant at the $p < .001$ level

consequence of the fact that SF is a compound of AS and $P^{\%}_{\text{Scored}}$. SF does not seem to be considerably better an indicator of IF than $M^{\text{AS}}_{\text{Scored}}$ or $P^{\%}_{\text{Scored}}$ alone.

Table 3 presents a hitlist: the 20 journals with the highest SF. Obviously, some of these journals are not mainstream Psychology journals. They are included in the list because they published papers that were related to psychology, however. The correlation between the SF and the IF among these 20 journals is .522 (p = .020).

Inspection of the highest ranked journals by SF, which indexes weighted social media coverage, shows that journals related to psychology are quite frequent with 5 journals among the top 20. Medical journals are also frequent (6), and we find general journals and journals related to biology and neurosciences. Many other fields are completely missing, however. Note, however, that papers were included only if they were related to the discipline of Psychology, to begin with. Given this, it is somewhat surprising that the dominance of journals containing the word "psychology" in their title is not more pronounced.

Table 3 Descriptives and Impact Factor of the 20 journals with highest Score Factor

Journal	P	P_{Scored}	$P^{\%}_{\text{Scored}}$	Score	$M^{\text{AS}}_{\text{Scored}}$	SD	Max	SF	IF
Science	284	255	89.8	16,635	65.2	107.6	1064	58.550	33.611
PloS Medicine	90	83	92.2	4127	49.7	71.9	421	45.823	14.429
Nature	276	140	50.7	8970	64.1	111.3	801	32.499	41.456
PloS Biology	59	54	91.5	1880	34.8	58.3	300	31.842	9.343
Psychological Science	555	457	82.3	16,970	37.1	66.4	771	30.533	4.940
Nature Reviews Neuroscience	59	54	91.5	1521	28.2	41.9	253	25.803	31.427
British Medical Journal	558	423	75.8	13,907	32.9	75.4	1016	24.938	17.445
Environmental Health Perspectives	49	29	59.2	1122	38.7	133.9	729	22.910	7.977
Clinical Infectious Diseases	29	14	48.3	657	46.9	135.4	506	22.653	8.886
Nutrition Journal	87	73	83.9	1949	26.7	59.0	372	22.401	2.597
Nature Neuroscience	212	176	83.0	4591	26.1	61.2	523	21.663	16.095
Cell	34	32	94.1	723	22.6	25.8	131	21.267	32.242
Perspectives on Psychological Science	113	87	77.0	2254	25.9	35.2	163	19.943	9.546
Pediatrics	484	321	66.3	8754	27.3	55.6	463	18.100	5.473
American Journal of Clinical Nutrition	187	137	73.3	3362	24.5	38.7	260	17.959	6.770
Trends in Cognitive Sciences	132	109	82.6	2301	21.1	35.7	269	17.429	21.965
Health Affairs	169	131	77.5	2904	22.2	32.9	164	17.205	4.966
Personality and Social Psychology Review	35	31	88.6	591	19.1	32.9	167	16.923	6.692
Journal of Personality and Social Psychology	304	229	75.3	5046	22.0	34.2	224	16.566	5.031
Biology Letters	140	104	74.3	2255	21.7	58.5	577	16.123	3.248

P = Number of papers, P_{Scored} = Number of papers with AS > 0, $P^{\%}_{\text{Scored}}$ Percentage of scored papers, Score = Accumulated AS, $P^{\%}_{\text{Scored}}$ Mean Score for all scored papers, SF Score Factor, IF Impact Factor, Ranking by SF

Articles

Most papers did not score at all in altmetrics. Indeed, only 57,087 of the 202,432 (28.2%) papers from WoS were mentioned in the online media at least once (AS > 0). A clear trend towards more attention from online media in recent years was noticeable: of the papers published in 2010 only 16.4% were mentioned in the online media at least once. This percentage was 26.1% in 2011, and was 41.1% in 2012 (see Table 4). Clearly, online media are becoming increasingly important as vehicles for disseminating scientific information in psychology.

At the article level, correlations between citations and the AS are not impressive, with a maximum of (log transformed) $r_{\log/\log} = .310$ in 2011. As expected, papers published earlier also had higher citations, with papers published in 2010 gaining almost thrice the citations of papers published in 2012. Indeed, publication year accounted for about 8% of the total variance in citations ($F(2, 57{,}084) = 2476$, $p < .001$, $R^2 = .080$). Interestingly, the AS was not nearly as influenced by publication year ($F(2, 54{,}784) = 132.3$, $p < .001$, $R^2 = .005$). This mirrors one of the basic differences between alternative and traditional metrics: alternative metrics are relative immediate and short-living, while most traditional metrics are delayed and cumulative. This different temporal dynamics puts a natural limit on the size of the correlation.

Average AS per paper scored was 8.4 (2.4 including articles never mentioned). Note, however, the highly skewed distribution of the AS: 39.8% of the papers mentioned online achieved a score of only 1, with 79.3% of articles scoring below the mean, while the 10 highest scoring articles (.02%, see Fig. 3 and Table 5) make up 2.0% of the total AS.

For the 10 highest scoring articles, contrary to the overall pattern, no significant relationship between citations and the AS was found ($r_{\log/\log} = -.303$, $p = .395$). Only one of the ten highest scoring articles for AS (# 7) is also found in the ten highest scoring articles for citations (ranked 2 out of all 57,087 articles; the second highest in citations in the top ten for AS (# 6) is ranked only 221 in citations).

Sources

The coverage for the various sources tracked by Altmetric.com varies vastly. An overview showing the results for the most important sources is given in Table 6. As can be seen, *Twitter* is by far the largest platform, covering 80% of all papers mentioned online. As such, in Twitter the correlation between AS and citations ($r_{\log/\log} = .096$, $p < .001$) is below the total correlation ($r_{\log/\log} = .196$, $p < .001$). Overall, Blogs are the best indicator

Table 4 Correlations and descriptives for the Altmetric Score and citations by year and total

Metric	N	$P^{\%}_{\text{Scored}}$	$P^{\%}_{\text{Scored}}$	M (SD)
AS^{2010}	10,322	16.4	.269***	9.1 (23.8)
CT^{2010}				32.2 (47.8)
AS^{2011}	18,398	26.1	.310***	8.3 (27.0)
CT^{2011}				20.8 (34.1)
AS^{2012}	28,367	41.1	.299***	8.2 (30.2)
CT^{2012}				12.9 (19.9)
$AS^{2010-2012}$	57,087	31.8	.302***	8.4 (28.1)
$CT^{2010-2012}$				18.9 (32.2)

AS Altmetric Score, *CT* Citations, $r_{\log/\log}$ log-transformed correlation

***$p < .001$

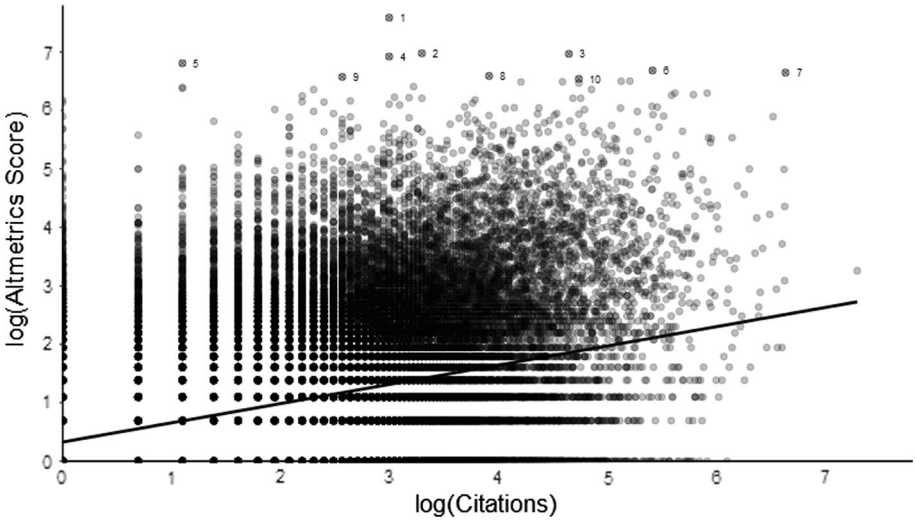

Fig. 3 Number of citations by AS, with regression line for individual papers. Axes are log-transformed. Papers with the 10 highest altmetric scores are numbered (see Table 5)

for citations, showing a medium correlation ($r_{\log/\log} = .258$, $p < .001$). Using Mendeley as a criteria instead of citations seems to yield even stronger results, as it averages a medium correlation with the total of all altmetric sources ($r_{\log/\log} = .272$, $p < .001$)

Discussion

This paper evaluates the relationship between traditional metrics and emerging alternative metrics. The source is published papers that are related to Psychology, published between 2010 and 2012. We extracted number of citations, and altmetric score (altmetrics.com) by June 2016, calculated various metrics, and evaluated their relationship. Out of a sample of nearly 250,000 papers about 240.000 were identified by a DOI. Of these, about 210,000 could be automatically allocated to journals with a discipline classification, and about 200,000 papers could be allocated to a journal with an impact factor. Among those, about 57,000 papers had an AS > 0. Note that all these papers have something to do with Psychology, as identified by the field "research area = Psychology" in Web of Science.

The main finding is that the relationship between traditional metrics and the AS, which measures the coverage of papers, journals, and disciplines in various alternative metrics depends on the level of analysis. An analysis in terms of different research fields (e.g. Biology, Economics and Business, Psychology and Cognitive Sciences) shows strong overlap: the correlation between citation counts and AS for 21 research fields was $r = .503$. This is impressive, showing that, in terms of entire research fields, traditional and alternative metrics measure similar things. At the level of 125 subfields the relationship was also strong ($r = .417$). However, there was considerable variability between fields and subfields, with correlations varying between $r = .106$ (Human Factors) and $r = .467$ (Communication and Textual Studies). However, the more fine-grained the level of analysis, the smaller the correlation: at the level of individual papers the correlation was only $r = .302$. This is partly due to the fact that, with aggregation, error variance gets

Table 5 Rank, Altmetric Score, and publication information for 10 highest scoring articles

Rank	AS	Citations	Title	Authors	Journal	Year
#1	1975	20	The Power of Kawaii: Viewing Cute Images Promotes a Careful Behavior and Narrows Attentional Focus	Nittono, Fukushima, Yano and Moriya	PLoS ONE	2012
#2	1073	27	Enclothed Cognition	Adam and Galinsky	Journal of Experimental Social Psychology	2012
#3	1064	104	Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips	Sparrow, Liu and Wegner	Science	2011
#4	1016	20	The Truth about Sports Drinks	Cohen	British Medical Journal	2012
#5	905	3	The Eyes Don't Have It: Lie Detection and Neuro-Linguistic Programming	Wiseman et al.	PloS ONE	2012
#6	801	224	A 61-Million-Person Experiment in Social Influence and Political Mobilization	Bond et al.	Nature	2012
#7	771	759	False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant	Simmons, Nelson and Simonsohn	Psychological Science	2011
#8	729	50	Developmental Fluoride Neurotoxicity: A Systematic Review and Meta-Analysis	Choi, Sun, Zhang and Grandjean	Environmental Health Perspectives	2012
#9	717	13	Creativity in the Wild: Improving Creative Reasoning through Immersion in Natural Settings	Atchley, Strayer, and Atchley	PloS ONE	2012
#10	693	114	Short-Term Music Training Enhances Verbal Intelligence and Executive Function	Moreno et al.	Psychological Science	2011

cancelled out. However, it also indicates that alternative metrics are an additional, and largely independent source of information at the level of individual papers. Highly cited papers may easily fail in the short-lived online world, and online star papers may fail in attracting citations. At the level of subfields and fields, or even disciplines (not investigated here), alternative metrics appear to offer less unique insights. Nevertheless, in terms of variance explained, it pays to consider these metrics even at these levels, since a correlation of $r = .50$ still explains only 25% of the variance.

To assess the importance of altmetrics on the journal level, we proposed a new metric: the Score Factor. The SF measures each journal's presence in the online media by combining coverage (whether a piece is mentioned at all in the social media) with density (how often a covered piece is mentioned). This makes sense, since most papers do not make it into an altmetric score. In addition, coverage in the online media is not restricted to the scientific community, although, in reality, tweets to scientific papers tend to come from educated individuals (with an over-representation of social and computer scientists, and

Table 6 Descriptives and correlations for sources from Altmetric.com

Source[a]	P_{Scored}	$P^{\%}_{\text{Scored}}$	$M^{\text{AS}}_{\text{Scored}}$	sd	$r_{\text{log/log}}$		
					Citations	Mendeley	CiteULike
Twitter	45,664	80.0	5.6	20.2	.096***	.206***	.194***
Facebook	10,932	19.1	2.4	7.0	.125***	.143***	.149***
Google^{+}	2909	5.1	2.2	7.9	.113***	.163***	.207***
Blogs	9591	16.8	1.8	2.1	.258***	.268***	.273***
News	5372	9.4	2.5	4.2	.217***	.204***	.166***
Other	21,519	37.7	2.7	23.5	.179***	.185***	.195***
Total	57,087	100	6.6	28.8	.196***	.272***	.245***

P_{Scored} = papers with AS > 0; $P^{\%}_{\text{Scored}}$ = percentage of scored papers for each source; $M^{\text{AS}}_{\text{Scored}}$ = mean score of all scored articles for each source

***$p < .001$

[a]Papers can be discussed in multiple sources

underrepresentation of mathematical, physical, and life scientists; Ke et al. 2017). In general, the SF offers information on a journal's importance in the online media different to the traditional Impact Factor. This is important since the journal impact factor might not be as central and harmless as it seems (Seglen 1997; Colquhoun 2003; Bollen et al. 2009). As the Editor-in-Chief of *Science* Bruce Alberts (2013) phrased it in an editorial on impact factor distortions: "The misuse of the journal impact factor is highly destructive, inviting a gaming of the metric that can bias journals against publishing important papers in fields (such as social sciences and ecology) that are much less cited than others (such as biomedicine). And it wastes the time of scientists by overloading highly cited journals such as Science with inappropriate submissions from researchers who are desperate to gain points from their evaluators" (p. 787).

We found a considerable correlation between SF and IF of about $r = .4$. Note, however, that correlations between the AS and citation frequency for articles which have been scored are small or even non-existent for the highest scoring papers. This indicates that, although a general relationship exists between alternative and traditional metrics, the relationship declines for individual papers and might easily be non-existent for important papers: what is relevant for the online community needs not be relevant to the scientific community. One of biggest skeptics of bibliometrics, and altmetrics in particular, Colquhoun (2014), in his blog, explains this occurrence as follows: "Scientific works get tweeted about mostly because they have titles that contain buzzwords, not because they represent great science". This notion was however, not confirmed by Taylor and Plume (2014), who examined highly shared papers using altmetric data. They were interested in examining whether articles attracting social media attention also are successful in getting the attention of scholars and the mass media. In their qualitative analysis of the top .5% of papers for activity in the social media they failed to find a bias for titillating or eye-catching keywords. Rather, their evaluation is more positive with respect to the scientific value. However, they found that most of the traffic in social media is related to summaries of research, rather than primary research articles themselves (but see Haustein et al. 2015, for a somewhat different result).

Although the distribution of scientific research in the online media has been on the rise over the years, impact still is unevenly distributed. Twitter is the largest platform (and presumably the only one that is genuinely relevant, as it is the only platform to reach a coverage above 20% for the distribution of scholarly publications and findings on the web, as is evident from several studies either relying on tweets as the measurement of alternative metrics (de Winter 2015), or evaluating the usage of internet platforms (Thelwall et al. 2013). This, in some sense, is good news, since Twitter is used mainly by non-academics. However, it seems that scientific material is mainly tweeted by scientists (Ke et al. 2017). Thus, the distribution of scientific material via Twitter among the public may be less than optimal.

As for psychology, comparable results to previous studies on other fields of research such as biomedicine (Haustein et al. 2014a) or astrophysics (Haustein et al. 2014b) are existing. The general picture is that correlations between altmetrics and citations are positive but small, indicating different roles of measuring scientific impact for traditional metrics and alternative metrics. Instead of dismissing those discrepancies as incompatible metrics, differing indicators should instead be used to create a framework for the concurrent use of various kinds of scientometric indicators to establish a more extensive assessment of the scientific impact of scholarly publications. Such a 'scholarly network' (Taylor 2013) could help to establish a more complete picture of scholarly impact, which at present is still missing (Priem et al. 2012). We want to add, however, that our findings imply that the AS is adequate for evaluating broad research areas, but should be used with caution for evaluating individual scholars, or individual papers. In addition, altmetrics are better seen as a complement rather than a substitute of traditional metrics like the impact factor. Substitution of traditional metrics, most notably of the impact factor may be desirable given a number of problems related to this traditional metrics (e.g., that the impact factor is negotiated, methodologically flawed, and irreproducible; see Brembs et al. 2013; Fernández-Delgado and Gómez 2015), but for the time being, alternative metrics, and the AS in particular, also suffer from serious limitations (Gumpenberger et al. 2016). The number of citations of a paper—not the impact factor of the journal that published the paper—might still be the best single indicator of a paper's quality. This number can, and will, increase over the years, while any alternative metric, because of its short half-life, will stagnate soon after publication. Thus, citations measure intermediate and long-term academic influence, while alternative metrics measure immediate academic and non-academic influence. Correlations will not be high under those circumstances.

The formation of a scholarly network by the involvement of scholars in the social media could furthermore establish a link between the scientific community and the public. This could help to involve the public in the scientific progress and would be a move from an exclusive scientific community to a truly overarching community with real time relevance. This concept is supported by the results from the MESUR Project (Bollen et al. 2007), indicating that usage-based metrics are indeed of value for the measurement of scholarly impact (Bollen et al. 2008).

All in all, alternative metrics are still on the verge of validation and have yet to prove themselves to be of any use for the scientific community. Most notably, care should be taken when linking them to an individual researcher's prestige. There are plenty of possibilities for the quantitative exploration of scientific publications, but any quantitative analysis should always bear in mind that scientific progress depends on the quality of papers, rather than on the prestige of outlets. Bear in mind Tressoldi et al's. (2013) answer to the question whether high impact equals high statistical standards: "not necessarily so" (p. e56180), they say. Whether the same ought to be said about alternative metrics is open

to debate. It is unlikely that a high AS would indicate high statistical standards, but one would hope that a high AS indicates high relevance of the scientific work for the academic and non-academic public. As the public gives the resources for research, papers with high AS succeed in reciprocity, giving something interesting back to the public.

Acknowledgements Open access funding provided by Paris Lodron University of Salzburg.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Alberts, B. (2013). Impact factor distortions. *Science, 340*(6134), 787. https://doi.org/10.1126/science.1240319.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to "webometrics". *Journal of Documentation, 53*(4), 404–426. https://doi.org/10.1108/EUM0000000007205.

Archambault, É., Beauchesne, O., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th international conference of the international society for scientometrics and informetrics* (pp. 66–77).

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social web. Retrieved from http://arxiv.org/abs/1205.5611.

Bollen, J., de Sompel, H. V., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE, 4*(6), e6022. https://doi.org/10.1371/journal.pone.0006022.

Bollen, J., Rodriguez, M. A., & Van de Sompel, H. (2007). *MESUR: Usage-based metrics of scholarly impact* (p. 474). New York: ACM Press. https://doi.org/10.1145/1255175.1255273.

Bollen, J., Van de Sompel, H., & Rodriguez, M. A. (2008). *Towards usage-based impact metrics: first results from the mesur project* (p. 231). New York: ACM Press. https://doi.org/10.1145/1378889.1378928.

Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience, 7,* 291. https://doi.org/10.3389/fnhum.2013.00291.

Brossard, D., & Scheufele, D. A. (2013). Science, new media, and the public. *Science, 339*(6115), 40–41. https://doi.org/10.1126/science.1232329.

Castells, M. (Ed.). (1999). *Critical education in the new information age*. Lanham: Rowman and Littlefield.

Coile, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of the American Society for Information Science, 28*(6), 366–370.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: University of Chicago Press.

Colquhoun, D. (2003). Challenging the tyranny of impact factors. *Nature, 423*(6939), 479. https://doi.org/10.1038/423479a.

Colquhoun, D., & Plested, A. (2014, January 16). Why you should ignore altmetrics and other bibliometric nightmares. Retrieved from http://www.dcscience.net/2014/01/16/why-you-should-ignore-altmetrics-and-other-bibliometric-nightmares/. Accessed July 2017.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology, 66*(10), 2003–2019. https://doi.org/10.1002/asi.23309.

de Winter, J. C. F. (2015). The relationship between tweets, citations, and article views for PLoS ONE articles. *Scientometrics, 102*(2), 1773–1779. https://doi.org/10.1007/s11192-014-1445-x.

Dennis, W. (1954). Productivity among American psychologists. *American Psychologist, 9*(5), 191–194.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research, 13*(4), e123. https://doi.org/10.2196/jmir.2012.

Fernández-Delgado, F. C., & Gómez, M. T. R. (2015). The index and the moon: Mortgaging scientific evaluation. *International Journal of Communication, 9,* 1880–1887.

Garfield, E. (1964). Science Citation Index: A new dimension in indexing. *Science, 144*(3619), 649–654.
Gumpenberger, C., Glänzel, W., & Gorraiz, J. (2016). The ecstasy and the agaony of the altmetric score. *Scientometrics, 108,* 977–982.
Haustein, S., Bowman, T. D., Holmberg, K., Peters, I., & Larivière, V. (2014a). Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. *Aslib Journal of Information Management, 66*(3), 279–296. https://doi.org/10.1108/AJIM-09-2013-0081.
Haustein, S., Costas, R., & Lariviere, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaborative patterns. *PLoS ONE, 10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495.
Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014b). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology, 65*(4), 656–669. https://doi.org/10.1002/asi.23101.
Ke, Q., Ahn, Y.-Y., & Sugimoto, C. R. (2017). A systematic identification and analysis of scientists on Twitter. *PLoS ONE, 12*(4), e0175368. https://doi.org/10.1371/journal.pone.0175368.
Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*, 575–603. https://doi.org/10.1007/s11192-010-0202-z.
Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(12), 317–323.
Nath, R., & Jackson, W. M. (1991). Productivity of management information systems researchers: Does Lotka's law apply? *Information Processing and Management, 27,* 203–209.
Patthi, B., Prasad, M., Gupta, R., Singla, A., Kumar, J. K., Dhama, K., et al. (2017). Altmetrics—A collated adjunct beyond citations for scholarly impact: A systematic review. *Journal of Clinical and Diagnostic Research, 11,* ZE16–ZE20. https://doi.org/10.7860/jcdr/2017/26153.10078.
Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday 15. http://firstmonday.org/htbin/cgiwrap/bin/. Accessed July 2017.
Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved from http://arxiv.org/abs/1203.4745.
Rushton, J. P. (1984). Evaluating research eminence in psychology: The construct validity of citation counts. *Bulletin of the British Psychological Society, 37,* 33–36.
Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ, 314*(7079), 497. https://doi.org/10.1136/bmj.314.7079.497.
Simonton, D. K. (1988a). *Scientific genius: A psychology of science*. Cambridge: Cambridge University Press.
Simonton, D. K. (1988b). Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin, 104*(2), 251–267.
Smith, R., & Fiedler, F. E. (1970). *The measurement of scholarly work in academic institutions: a critical review of the literature [by] Richard Smith and Fred E. Fielder*. Seattle: University of Washington.
Taylor, M. (2013). Towards a common model of citation: Some thoughts on merging altmetrics and bibliometrics. *Research Trends, 35,* 19–22.
Taylor, M., & Plume, A. (2014). Party papers or policy discussions: An examination of highly shared papers using altmetric data. *Research Trends, 36,* 17–20.
Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE, 8*(5), e64841. https://doi.org/10.1371/journal.pone.0064841.
Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology, 39,* 81–135. https://doi.org/10.1002/aris.1440390110.
Tressoldi, P. E., Giofré, D., Sella, F., & Cumming, G. (2013). High impact = high statistical standards? Not necessarily so. *PLoS ONE, 8*(2), e56180. https://doi.org/10.1371/journal.pone.0056180.
Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. *Scientometrics, 101*(2), 1491–1513. https://doi.org/10.1007/s11192-014-1264-0.
Zhang, J., Ning, Z., Bai, X., Kong, X., Zhou, J., & Xia, F. (2017). Exploring time factors in measuring the scientific impact of scholars. *Scientometrics, 112*(3), 1301–1321. https://doi.org/10.1007/s11192-017-2458-z.